A fast and robust method for the identification of face landmarks in profile images

A Fast and Robust Method for the Identification of Face Landmarks in Profile Images

ANDREA BOTTINO, SANDRO CUMANI
Dipartimento di Automatica e Informatica
Politecnico di Torino
Corso Duca degli Abruzzi, 24 – 10129 Torino
ITALY
andrea.bottino@polito.it http://www.polito.it/cgvg

Abstract: - Several applications in Computer Vision, like recognition, identification, automatic 3D modeling and animation and non conventional human computer interaction require the precise identification of landmark points in facial images. Here we present a fast and robust algorithm capable of identifying a specific set of landmarks on face profile images. First, the face is automatically segmented from the image. Then, the face landmarks are extracted. The algorithm is based on the local curvature of the profile contour and on the local analysis of the face features. The robustness of the presented approach is demonstrated by a set of experiments were ground truth data are compared with the result of our algorithm. A percentage of 92% correct identification and a mean error of 3.5 pixels demonstrate the robustness of the approach, which is of paramount importance for several applications.

Key-Words: - silhouettes, profile images, face landmark, robust identification

1 Introduction

The problem of automatically identifying the precise location of face landmarks is an active research topic in computer vision and it has several interesting applications. For recognition, face landmarks can be used to derive a characteristic vector, based for instance on the distances between the feature points. In this way, faces can be represented by points in the reduced dimensionality space of the characteristic vectors, where point comparison can be used for recognition/identification ([2], [3], [4]). For these applications, precision in the location of the landmarks is critical. Another typical application is the reconstruction of 3D head model from single or multiple 2D images ([5], [6], [7]). The usual approach is to identify face landmarks and then use their location to constraint the transformation of a deformable 3D head model. Other interesting applications include motion capture [14], human computer interaction [15], teleconferencing [16], entertainment, animation [17] and many others.

The rationale of this work is to propose an automatic and robust approach able to identify and segment face features in profile images. This is part of a broader project aiming at developing an automatic tool for planning plastic surgery procedures. The tool suggests the possible outcome of a plastic surgery operation proposing a new quantitative approach able to automatically suggest effective patient-specific improvements of facial attractiveness [18].

Several approaches have been presented for face landmarks identification. In [5] the profile and frontal images are analyzed and landmarks are extracted exploiting the curvature of the profile and applying deformable templates for face features like mouth, nose and eyes. Multi-modal deformable templates are also used in [8]. More complex approaches use Gabor filters [9], neural networks [10], PCA [11], 2D discrete cosine transform [12], Fisher's linear discriminant [13].

In this work, we present a robust approach, based on simple and fast computer vision algorithms. The image is first segmented in order to extract the silhouette of the person's face. The feature points on the external contour of the silhouette are then extracted on the basis of the local curvature of the profile and on some geometric properties of its points. Then, the landmarks inside the face region are

identified analyzing in details the local characteristics of the image.
The paper is organized as follows. Section 2 describes the set of face landmarks we are trying to identify. The proposed approach is detailed in Section 3. The results obtained and future work are discussed in Sections 4 and 5 respectively.

2 Processing face profiles

The set of facial landmarks considered in this work is the following:

- Pronasale, (PN), the tip of the nose
- Nasion (NA), the superior extremity of the nose line, in the depressed area directly between the eyes
- Subnasal (SN) , the inferior extremity of the nose line
- Labrale Superius (LS) and Inferius (LI), on the upper and lower lip
- Cheilion (CH), located at the angle (corner) of the mouth
- Gnathion (GN), the inferior point of the mandible
- Exocanthion (EC), the outer corner of the eye
- Alare (AL) the outermost point of the nose wing

The names have been taken from the medical terminology [1]. In Fig. 1 the location of these landmark points is shown on a sample face.

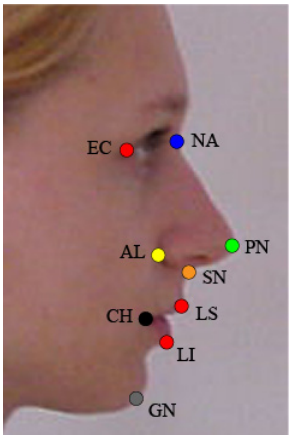

Fig. 1: the set of face landmarks to identify

3 Landmark identification

The algorithm works on images containing a single face and with uniform background color. Before identifying the face landmarks, some pre-processing is needed.
The first operation required is to extract the silhouette of the profile from the incoming image. Segmentation is performed through background subtraction. For this task, we require the images to have a background of an (almost) uniform color. First, we transform the image into the HSL color space. Then we compute the histogram of the H plane, containing the chromaticity information. The two main peaks correspond to skin-like and background pixels. An example can be seen in Fig. 2, where an image and the histogram of its hue plane are shown.

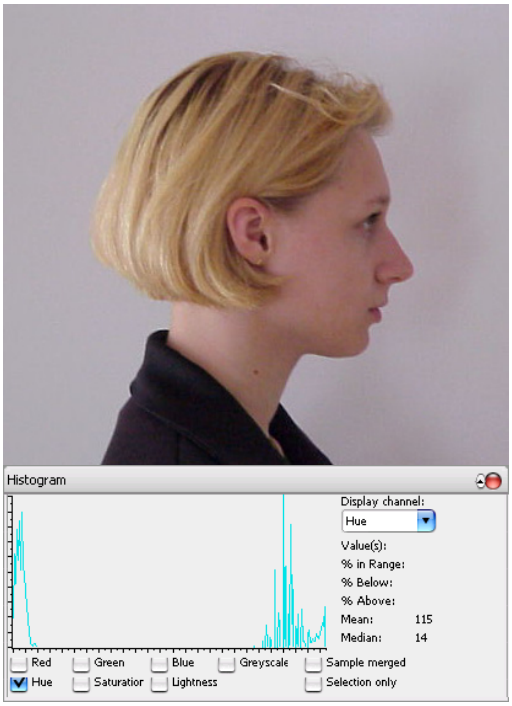

Fig. 2: an image and its hue histogram

Reasonably, the number of background pixels is greater than skin-like pixels. However, since we cannot rely on this assumption to identify the background color, we can consider the fact the skin has peculiar chromatic characteristics. It has been demonstrated ([19]) that different color models are not needed for different races of people. In other

words, except for albinos, the chromatic characteristics of the skin are similar for all races. This fact allows to build a priori a skin model that can be used to identify the peak corresponding to face pixels and, accordingly, to background pixels. The background model therefore is a simple interval around the background peak. In order to identify foreground pixels correctly, we must consider a limitation of the HSL model. When the brightness of a pixel is low, L close to 0, also S is low and the value of H becomes quite noisy. On the contrary, at very low saturations, hue is not defined. Therefore we must ignore pixels belonging to these two cases.
In order to remove clothes, we consider the skin-like pixels in the lower part of the image to define an "end of the neck" cut. Other non skin regions, for instance hairs, are not removed. Finally, median filtering and morphological operators are applied for noise and outliers removal, and the holes of the resulting connected components are removed with a flood fill algorithm. An example of the segmentation result is shown in Fig. 3. Several other techniques for face segmentation can be found in the literature ([22], [23], [24]).

Fig. 3: an image (left) and the corresponding head silhouette

The second operation required is to establish an orthogonal reference coordinate system UV aligned with the head. This can be done exploiting the image moments [20]. The moment of order $(p+q)$ $p,q=0,1,2\ldots$, for an intensity image I is defined as:

$$M_{pq} = \sum_x \sum_y x^p y^q I(x,y)$$

The centroid of R is:

$$C = \{\bar{x}, \bar{y}\} = \{M_{10}/M_{00}, M_{01}/M_{00}\}$$

The central moments of the image are defined as:

$$\mu_{pq} = \sum_x \sum_y (x-\bar{x})^p (y-\bar{y})^q I(x,y)$$

The orientation of the head can be deduced from the covariance ellipse built on the second order central moments of I. The covariance matrix is:

$$\mathrm{cov}[I(x,y)] = \begin{bmatrix} \mu'_{20} & \mu'_{11} \\ \mu'_{11} & \mu'_{02} \end{bmatrix}$$

where:

$$\mu'_{11} = \mu_{11} / \mu_{00}$$
$$\mu'_{20} = \mu_{20} / \mu_{00}$$
$$\mu'_{02} = \mu_{02} / \mu_{00}$$

The eigenvectors of this matrix correspond to the major and minor axes of the intensity image, so the orientation can thus be extracted from the angle between the coordinate axis and the eigenvector associated with the largest eigenvalue. This angle, θ, is given by:

$$\theta = \tfrac{1}{2} \arctan\left(\frac{2\mu'_{11}}{\mu'_{20} - \mu'_{02}} \right)$$

The U coordinate axis is the semi axis of the covariance ellipse pointing rightward in the image, while V is the one pointing upward (Fig. 4).

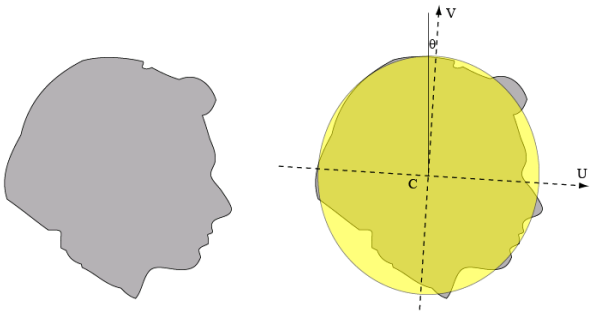

Fig. 4: the head silhouette (left) and the covariance ellipse built on the image moments (right)

The landmarks of interest can be substantially divided in two subgroups: the points lying on the outer contour of the face profile (NA, PN, SN, LS, LI, and GN) and the ones lying inside the face region (CH, EC, AC). The following subsections detail the operations required to locate the two subsets of landmarks.

3.1 Landmarks lying on the outer contour of the face profile

These points are lying on the frontier of R, and their common characteristic is that their curvature is a local extreme on the profile curve. For the sake of simplicity, we suppose that all faces are oriented towards the same direction (e.g. towards right); therefore, all the points of this set are lying on the half plane U>0.

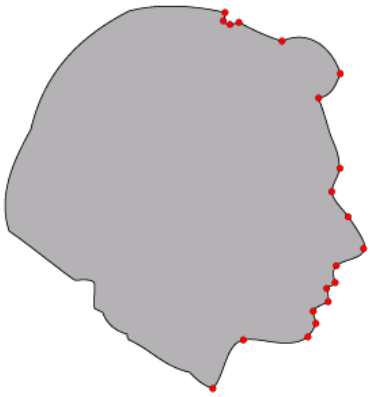

Fig. 5: the set of candidate points

Accurate radius of curvature for each point of the profile can be computed directly from the first and second order derivatives, evaluated by convolution with isotropic Gaussian derivative kernels [21]. The set of *candidate points* contains the local extremes of the curvature in the half plane U>0 (see Fig. 5 for an example).

The landmarks on the face profile belong to the candidate set and can be identified with a suitable set of rules. These rules are the following:

- The Pronasale (PN) is the point with greater U in the region U>0 and $V<t_{PN}$. The threshold t_{PN} is to avoid incorrect identification of the hairs as the Pronasale, and has been evaluated experimentally as 30% of the greater V of the candidate set (Fig. 6)
- The point P below the Pronasale ($V_P < V_{PN}$) at maximal distance from the center C, and whose ratio between P-C and PN-C distances does not fall below a threshold r_{GN}, is labeled as the Gnathion (GN). The threshold has been evaluated experimentally to 0.8
- The point P above the Pronasale ($V_P > V_{PN}$) and closer to the center, is identified as the Nasion (NA)
- The remaining points (SN, LS, LI) can be identified scoring the candidate points enclosed by PN and GN by means of their curvature sign, their distance with the line connecting PN and GN and their symmetry along the V axis, where the constraint $V_{LI}<V_{LS}<V_{SN}$ must be satisfied (see Fig. 7). During this step we also identify on the profile the lips intersection point, LL, as the point at highest curvature between LI and LS. This point will be used to identify the Cheilion

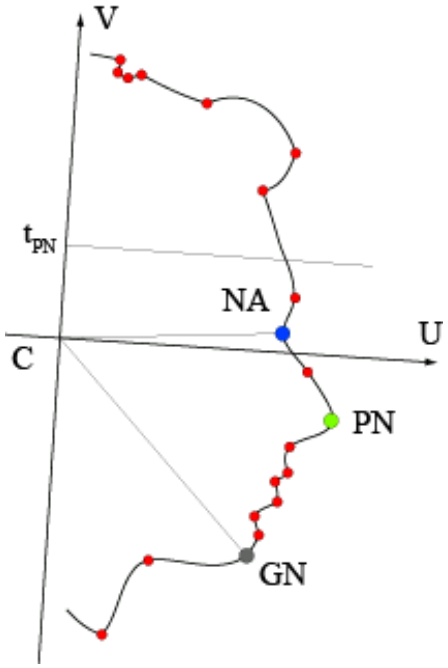

Fig. 6: identification of Pronasale, Nasion and Gnathion

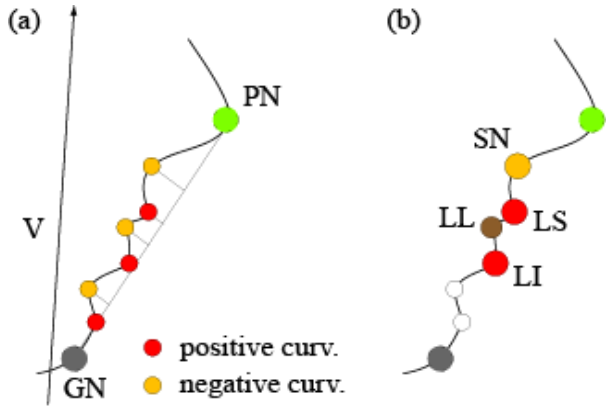

Fig. 7: identification of Subnasal, Labrale Superius and Inferius

3.2 Landmarks lying inside the face region

The identification of this subset of points requires an analysis of the face texture, involving the extraction of several image clues which are combined according to the specific feature.

In order to detail the steps of the algorithm, the following notation will be used. Let R^b_a be the orthogonal coordinate system centred in a whose u versor is parallel to b-a segment, and let $[u^b_a]_p$ and $[v^b_a]_p$ be the coordinates of point p in this system.

3.2.1 Exocanthion

The region were the EC can be located is first roughly identified by means of geometrical relationships involving the position of C, Pronasale and Nasion. This region is selected as Region of Interest (ROI_{EC}), where further processing is involved in order to reduce the computational burden. In the ROI_{EC}, we first determine strong corners of the image using eigenvalue analysis ([22]). The initial results are scored in order to obtain a set of four candidates. Scoring is a function of the projection of the points on the NA-PN and NA-GN segments.

Finally, suitable 5x5 filter masks are applied to ROI_{EC} as upper-lower eyelid and bottom-top left corner detectors. The masks are the following:

$$F_{up\_eyl} = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & -1 & -1 \\ 0 & 0 & -1 & 0 & 0 \\ -1 & -1 & 0 & 0 & 0 \end{bmatrix}$$

$$F_{low\_eyl} = \begin{bmatrix} -1 & -1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ 1 & 1 & 0 & -1 & -1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

$$F_{BL\_corner} = \begin{bmatrix} -1 & -1 & 1 & 2 & 2 \\ -1 & -1 & 1 & 2 & 2 \\ -1 & -1 & 0 & 0 & 0 \\ -1 & -1 & 0 & 0 & 0 \\ -1 & -1 & 0 & 0 & 0 \end{bmatrix}$$

$$F_{TL\_corner} = \begin{bmatrix} -1 & -1 & 0 & 0 & 0 \\ -1 & -1 & 0 & 0 & 0 \\ -1 & -1 & 0 & 0 & 0 \\ -1 & -1 & 1 & 2 & 2 \\ -1 & -1 & 1 & 2 & 2 \end{bmatrix}$$

The candidate point summing up the greater value from the four filters is identified as the Exocanthion. In Fig. 8, the original image (a), the ROI_{EC} and the four EC candidates (b), a false colour image showing the contributes of the two eyelid detectors (c) and the identified EC location (d) are shown.

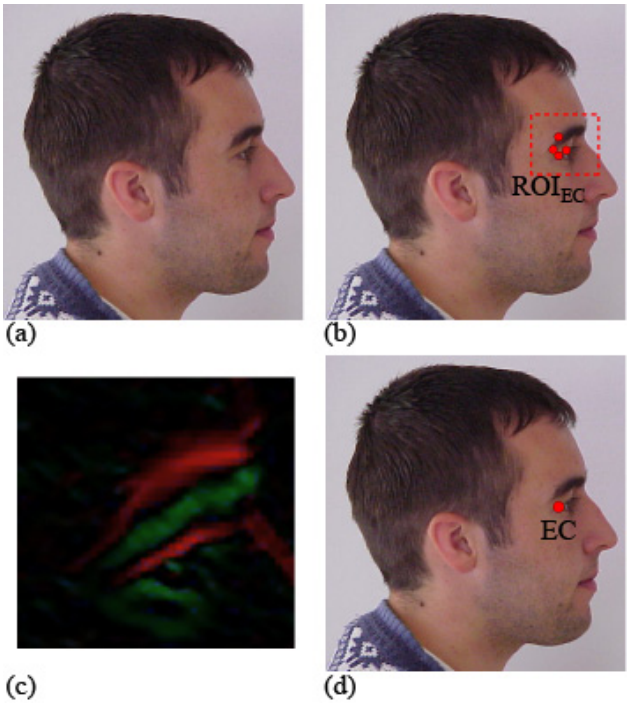

Fig. 8: original image (a), the four EC candidates (b), result of the eyelid detectors (c), final EC location (d)

3.2.2 Cheilion

Having defined a suitable ROI_{CH} combining the position of LS, LI and C, CH is usually one of the leftmost detected corners (that is, the one closer to C). The strong corners are identified by first applying a Canny edge detector, and then performing eigenvalue analysis on the resulting image. Since CH should also be near the lips intersection point on the profile (LL), whose position has been computed during the previous steps, the corners are scored combining their distance from the Labrale Inferior-Labrale Superior line and the distance of their projection with the projection of LL on the LI-LS line. The score of the candidate x_i is evaluated as:

$$[v_{LS}^{LI}]_{x_i}^2 - ([u_{LS}^{LI}]_{LL} - [u_{LS}^{LI}]_{x_i})^2$$

The corner obtaining the best score is selected as the Cheilion.

3.2.3 Alare

To identify the Alare, we start from identifying the image region to process. This region, the ROI_{AL}, is centered on an initial estimate, p_{AL}, of the position of the Alare, given by the intersection of the two segments Exocanthion-Labrale Superius and Cheilion-Nasion (Fig. 9(a)). The size of ROI_{AL} is proportional to the bounding box including Cheilion, Exocanthion and Pronasale. The resize factor, evaluated experimentally, is 0.4. In this region we try to identify the nose wing, filtering the intensity image (Fig. 9(b)) with the following mask:

$$F_{Nose\_Wing} = \begin{bmatrix} 1 & 0 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 \\ 2 & 0 & -2 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & -1 \end{bmatrix}$$

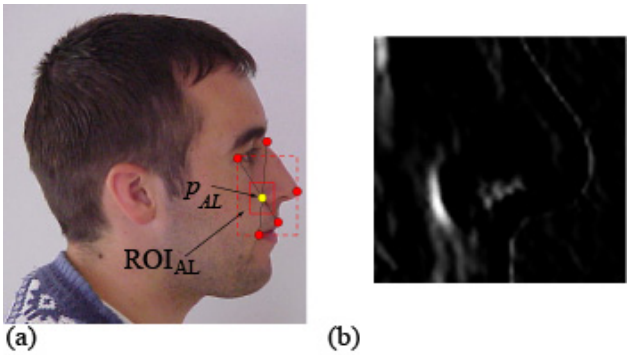

Fig. 9: ROI_{AL} (a) and output of the nose wing detector (b)

Finally, the strong corners in ROI_{AL} are identified and scored according to the intensity of the filtered image, to their distance with p_{AL} and to the distance obtained projecting the corner and the Pronasale on the line Nasion-Labial Superius. The rationale is that the Alare is usually at (almost) the same height as the Pronasale, and using the Nasion-Labrale Superior line it is possible to use this observation also for slightly rotated silhouettes. Given $g(x)$, the image filtered with F_{Nose_Wing}, the score of candidate x_i is given by:

$$g(x_i)\frac{1}{d_i+1}\frac{1}{\left|[u_{LS}^{NA}]_{PN}-[u_{LS}^{NA}]_{x_i}\right|+1}, \quad d_i=\left\|p_{AL}-x_i\right\|$$

The corner obtaining the best score is identified as the Alare. In case no corners are identified, the Alare candidate is chosen.

4 Experimental results

In order to validate the proposed algorithm, we have compared its results with ground truth data, obtained identifying manually the landmarks on a set of 72 images. The set of profiles is taken from the CVL Face Database of the University of Ljubljana [23]. All samples are 640x480 color images.

Landmark	*Mean error*
NA	3,98
PN	2,04
SN	1,72
LS	2,28
LI	1,84
GN	10,41
CH	1,91
EC	3,38
AC	4,42
Total	*3,55*

Table 1: mean identification errors in pixel

Landmark	*Points*	*Correct*	*Incorrect*	*Incorrect (%)*
NA	72	67	5	6,94%
PN	72	72	0	0,00%
SN	72	72	0	0,00%
LS	72	70	2	2,78%
LI	72	72	0	0,00%
GN	72	40	32	44,44%
CH	72	72	0	0,00%
EC	72	67	5	6,94%
AC	72	68	4	5,56%
Total	*648*	*600*	*48*	*7,41%*

Table 2: percentage of correct identifications

In Table 1 the mean identification error in pixels is listed. As it can be seen, excluding the case of Gnathion that will be discussed later, the mean error is below 4.5 pixels, a number that is comparable with the human mistake in evaluating the ground truth data. The results are also given in Table 2 in terms of correct/incorrect identification, where identification is correct if its distance in pixel from the ground truth location is lower then a predefined threshold (in this case, the threshold has been set to six). The results are given separately for each landmark, and globally. Overall, the resulting percentage of 92% correct identifications is a very good result, and demonstrates the robustness of the algorithm. However, as it can be seen analyzing the results in details, some of the landmarks have a ratio of correct identification which is not very satisfactory. This is for instance the case of Gnathion, whose identification ratio is about 56%. This result is influenced by two factors. First, the chin can contain distracters, like beard or hairs, which produce a local extreme of the curvature in their correspondence. Second, in several cases it is difficult to locate precisely the position of the Gnathion, and its location is questionable, especially when the chin has an almost uniform curvature (see Fig. 10). Regarding the Nasion, its relatively high error rate is mainly due to small errors in the background segmentation. In general, we can say that the profile curvature is a very powerful tool for identifying landmarks, while other image clues are less robust. The last fact is also related to the resolution and quality of the incoming images.

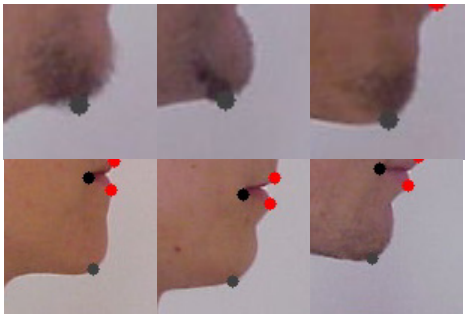

Fig. 10: incorrect Gnathion identification

In order to improve the current algorithm, better identification rules must be defined for those landmarks not reaching a satisfactory result (such as Gnathion, Exocanthion and Nasion).
As for the computational manageability, the mean execution time for the 72 images is 0.43 s on an Intel Core2 Duo T7700, 2.40Ghz and 2GB of RAM. This proves that the algorithm, besides being robust, is also fast in terms of execution times.

5 Conclusions

In this paper we have presented a fast and robust algorithm for identifying face landmarks on profile images. First, the approach extracts the silhouette of the face from the image by background subtraction. The curvature of the points on the silhouette boundary is computed in order to select a set of candidates for the subset of landmarks located on the profile. These are identified by means of simple geometric rules, based on the position of the center of the face, on its orientation and on the relative position of the candidates. The landmarks inside the face are then identified applying several filters and detecting the edges of the image.
The algorithm has been tested on a public database in order to validate its results. The experiments show that the proposed approach is very precise, with a percentage of correct identification of the 92%, and very fast, with an overall average execution time of 0.43 s. It should be outlined, however, that not all the results are satisfactory. For instance, the Gnathion is identified incorrectly in almost half of the available cases. Even if deciding its real location is questionable, this result is an indication that a different approach must be studied for this landmark.
The algorithm has been developed as part of a more complex system for planning plastic surgery operations. A preliminary version of this system has been developed for 2D profile images. Currently, we are planning to extend the approach to 3D and, regarding the topic discussed in this paper, further work will involve the automatic identification of face landmarks on 3D scanned head models.

References

[1] John C. Kolar, Elizabeth M. Salter, *Craniofacial Anthropometry: Practical Measurement of the Head and Face for Clinical, Surgical,and Research Use*. Charles C. Thomas Publisher, 1997

[2] Shi J., Samal A., Marx D.; "*How effective are landmarks and their geometry for face recognition?*", Computer vision and image understanding, 2006, vol. 102, no2, pp. 117-133

[3] Jiazheng Shi; Samal, A.; Marx, D.; "*Face recognition using landmark-based bidimensional regression*", Fifth IEEE International Conference on Data Mining, 7-30 Nov. 2005

[4] Beumer G.M., Tao Q., Bazen A.M., Veldhuis R.N. *A Landmark Paper in Face Recognition*. Proc. Int. Conf. on Automatic Face and Gesture 2006, pp. 73-78.

[5] Y.X. Hu, D.L. Jiang, S.C. Yan, H.J. Zhang, "*Automatic 3D Reconstruction for Face Recognition*", Proceedings of FGR2004

[6] Changhu Wang; Shuicheng Yan; Hongjiang Zhang; Weiying Ma, "*Realistic 3D Face Modeling by Fusing Multiple 2D Images*". Proc. Multimedia Modelling Conference, 2005, pp 139 – 146

[7] Ming-Shing Su, Chun-Yen Chen, Kuo-Young Cheng, "*An Automatic Construction of a Person's Face Model from the Person's Two Orthogonal Views*," GMP'02, p. 179, 2002

[8] Bottino A., "*Real time head and facial features tracking from uncalibrated monocular views*", Proc. ACCV 2002

[9] Li B., Chellappa R.," *Gabor attributes tracking for face verification*", Proc. International Conference on Image Processing, 2000. pp. 45-48

[10] Reinders M.J.T., Koch R.W.C., Gerbrands J.J., "*Locating facial features in image sequences using neural networks*". Proc. Automatic Face and Gesture Recognition, 1996,pp. 230 – 235

[11] M. Turk, A. Pentland, "*Eigenfaces for recognition*", J. Cognitive Neurosci. 3 (1) (1991) 71–86

[12] S. Eickler, S. M Wuller, G. Rigoll, "*Recognition of JPEG compressed face images based on statistical methods*", Image Vision Comput. 18 (4) (2000) 279–287

[13] P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, "*Eigenfaces vs. Fisherfaces: recognition using class specific linear projection*", IEEE Trans. Pattern Anal. Mach. Intell. 19 (7) (1997) 711–720

[14] Yunshu Hou, Hichem Sahli, Ravyse Ilse, Yanning Zhang, Rongchun Zhao, "*Robust Shape-Based Head Tracking*", Advanced Concepts for Intelligent Vision Systems, 2007, pp. 340-351

[15] C. Manresa-Yee, J. Varona, F.J. Perales. "*Face-Based Perceptual Interface for Computer-Human interaction*", Proceedings WSCG'06, Plzen, February 2006

[16] D. Machin, "*Real-Time Facial Analysis for Virtual Teleconferencing*", Proc. International Conference on Automatic Face and Gesture Recognition

[17] Shuo Wang, Xiaocao Xiong, Yan Xu, Chao Wang, Weiwei Zhang, Xiaofeng Dai, Dongmei Zhang, "*Face-tracking as an augmented input in video games: enhancing presence, role-playing and control*", Proceedings of the SIGCHI conference on Human Factors in computing systems, 2006, Pages: 1097 – 1106

[18] A. Bottino, A. Laurentini, L. Rosano, "*A New Computer-aided Technique for Planning the Aesthetic Outcome of Plastic Surgery*", Proceedings of 16th Int. Conf. on Computer Graphic, Visualization and Computer Vision WSCG 2008, Plzen (Cz), 4-7 February, 2008

[19] G. R. Bradski. *Computer video face tracking for use in a perceptual user interface*. Intel Technology Journal, Q2 1998

[20] M. K. Hu, *Visual Pattern Recognition by Moment Invariants*, IRE Trans. Info. Theory, vol. IT-8, pp.179-187, 1962

[21] C. van Wijk, R. Truyen, R. E. van Gelder, L. J. van Vliet, F. Vos, *On Normalized Convolution to Measure Curvature Features for automatic Polyp Detection*, proceedings MICCAI, 2004, pp. 200-208

[22] Baek et al., "Face Region Detection Using DCT and Homomorphic Filter", Proceedings of the 6th WSEAS International Conference on Signal Processing, Robotics and Automation, Corfu Island, Greece, February 16-19, 2007

[23] Saaidia et al., "Face detection by neural network trained with Zernike moments", Proceedings of the 6th WSEAS International Conference on Signal Processing, Robotics and Automation, Corfu Island, Greece, February 16-19, 2007

[24] J. Zapata and R. Ruiz, "A Hybrid Snake for Selective Contour Detection", Proceedings of the 6th WSEAS International Conference on Signal Processing, Robotics and Automation, Corfu Island, Greece, February 16-19, 2007

[25] J. Shi and C. Tomasi. *Good Features to Track*. IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WA, 1994

[26] CVL FACE DATABASE: Computer Vision Laboratory, University of Ljubljana, Slovenia, http://www.lrv.fri.uni-ljsi/facedb.html

Fig. 11